A scalable computational framework for establishing long-term behavior of stochastic reaction networks*

Corentin Briat†, Ankit Gupta‡, and Mustafa Khammash§

Department of Biosystems Science and Engineering (D-BSSE), Swiss Federal Institute of Technology–Zürich, 4058 Basel, Switzerland

April 22, 2013

Abstract

Reaction networks are systems in which the populations of a finite number of species evolve through predefined interactions. Such networks are found as modeling tools in many disciplines, spanning biochemistry, epidemiology, pharmacology, ecology and social networks. It is now well-established that, for small population sizes, stochastic models for reaction networks are necessary to capture *randomness* in the interactions. The tools for analyzing them, however, still lag far behind their deterministic counterparts. In this paper, we bridge this gap by developing a constructive framework for examining the long-term behavior and stability properties of the reaction dynamics in a stochastic setting. In particular, we address the problems of determining *ergodicity* of the reaction dynamics, which is analogous to having a globally attracting fixed point for deterministic dynamics, and determining *moment bounds* for the underlying stochastic process. Theoretical and computational solutions for these problems are obtained by utilizing a blend of ideas and techniques from probability theory, linear algebra, polynomial analysis and optimization theory. We demonstrate that stability properties of a wide class of networks can be assessed from theoretical results that can be recast as *efficient* and *scalable* linear programs, well-known for their tractability. It is notably shown that the computational complexity is often linear in the number of species, but worst-case quadratic. We illustrate the validity, the efficiency and the universality of our results on several reaction networks arising in fields such as biochemistry, epidemiology and ecology.

1 Introduction

Reaction networks represent a modeling paradigm that finds applications in many areas of science. Examples include, chemical reaction networks [9], cell signalling networks [28], gene expression networks [32], metabolic networks [30], pharmacological networks [3], epidemiological networks [17] and ecological networks [2]. Traditionally, reaction networks are mathematically analyzed by expressing the dynamics as a set of ordinary differential equations. Such a deterministic model is reasonably accurate when the number of network participants is *large*. However, when this is not the case, the discrete nature of the interactions becomes important and the dynamics is inherently *noisy*. This *random* component of the dynamics cannot be ignored as it can have a significant impact on the macroscopic properties of the system [15, 24, 23]. To account for this randomness and study its effects, a stochastic formulation of the dynamics is necessary. The most common approach is to model the dynamics as a continuous-time Markov process whose states denote the current population size. Many recent articles have been devoted to stochastic models and their analysis in view of understanding the role of noise and its impact on the system's behavior [8, 1, 22, 25].

Among the two modeling approaches, the deterministic approach is far more tractable than the stochastic approach. This is not surprising since our knowledge of ordinary differential equations is fairly advanced. For deterministic models, many tools allowing us to analyze the system without simulating the trajectories are indeed available. Unfortunately, this is not the case for stochastic models. Most papers that use such models rely on heavy computational techniques such as the simulation of several trajectories in order to determine the relevant characteristics of the system. More importantly, since one can only simulate a finite number of trajectories for a finite amount of time, properties like long-term behavior and stability cannot be verified through such simulations. In this paper, we overcome this problem and find a direct way to examine such properties for the stochastic model, without relying on simulations. Our approach combines stochastic analysis along with linear algebra, polynomial analysis and optimization techniques. For the latter we restrict ourselves to Linear Programming in this paper. Refinements of our method using more advanced optimization techniques will be considered elsewhere.

2 Reaction networks

Before we introduce the properties we study in this paper, we formally describe our reaction network. Moti-

*Corentin Briat and Ankit Gupta contributed equally to this work

†mail: corentin.briat@bsse.ethz.ch; url: http://www.briat.info, http://www.bsse.ethz.ch/ctsb/people/postdoc/briat

‡mail: ankit.gupta@bsse.ethz.ch; url: http://www.bsse.ethz.ch/ctsb/people/postdoc/gupta

§mail: mustafa.khammash@bsse.ethz.ch; url: http://www.bsse.ethz.ch/research/Professors/khammash\protect_cv

vated by the literature on chemical kinetics, we refer to the network participants as *molecules* which may belong to one of d species. There are K reactions in the network and for any $k = 1, \ldots, K$, the *stoichiometric* vector $\zeta_k = (\zeta_{k,1}, \ldots, \zeta_{k,d})$ denotes the change in the number of molecules in each of the species due to the k-th reaction. When the *state* of the system is x, the k-th reaction fires at rate $\lambda_k(x)$. The functions $\lambda_1, \ldots, \lambda_K$ are known as the *propensity* functions in the literature. The notion of *state* is different for deterministic and stochastic approaches. In the deterministic setting, the state is a vector of *concentrations* of the d species, while in the stochastic setting, the state refers to the vector of molecular counts of the d species.

2.1 Deterministic models

Consider the deterministic model for the reaction network described above. If the initial state is x_0, then the evolution of concentrations is given by $(\phi_{x_0}(t))_{t\geq 0}$ which satisfies the Reaction Rate Equations (RRE) of the form

$$\frac{d\phi_{x_0}(t)}{dt} = \sum_{k=1}^{K} \lambda_k(\phi_{x_0}(t))\zeta_k \quad \text{with} \quad \phi_{x_0}(0) = x_0. \tag{1}$$

We are interested in the long-term behavior and stability of our reaction dynamics. More precisely, we would like to check if the following conditions are satisfied.

DC1 For any x_0, there is a compact set $\mathcal{K}(x_0)$ such that $\phi_{x_0}(t) \in \mathcal{K}(x_0)$ for all $t \geq 0$.

DC2 There exists a compact set $\mathcal{K}_0$ such that for any x_0, we have $\phi_{x_0}(t) \in \mathcal{K}_0$ for large values of t.

DC3 There is a x_{eq} such that for any x_0 we have $\phi_{x_0}(t) \to x_{\text{eq}}$ as $t \to \infty$.

The first condition, **DC1**, says that for any x_0, the entire trajectory $(\phi_{x_0}(t))_{t\geq 0}$ stays within some compact set. We would expect this to be true for most *realistic* systems. Hence a violation of this property may suggest a *flaw* in the deterministic model. The second condition, **DC2**, says that there is an *attractor* set for the dynamics, where all the trajectories eventually lie, irrespective of their starting point. This is related to the notion of *permanence* for reaction networks. The last condition, **DC3**, says that there is a globally attracting fixed point for the deterministic model. Given a reaction network one can verify these conditions by using techniques from the theory of dynamical systems [21, 31], without having to simulate the paths of the deterministic process. There is also a general theory available that aims to check condition **DC3** for reaction networks satisfying mass-action kinetics (see [11, 12, 19, 18]). Our goal in this paper is to develop a theoretical and computational framework for verifying conditions similar to **DC1**, **DC2** and **DC3** for stochastic models of reaction networks.

2.2 Stochastic models

Consider the stochastic model corresponding to the reaction network described above. In this setting the firing of reactions are discrete events. When the state of the system is x, the k-th reaction fires after a time which is an exponentially distributed random variable with rate $\lambda_k(x)$. The dynamics can be represented by the Markov process $(X_{x_0}(t))_{t\geq 0}$ where x_0 is the initial state. Note that if $X_{x_0}(t) = (X_1(t), \ldots, X_d(t))$, then $X_i(t)$ is the *number* of molecules of the i-th species at time t. It is important to choose a suitable state space $\mathcal{S}$ for this Markov process. If $\mathbb{N}_0$ is the space of non-negative integers then let $\mathcal{S}$ be the smallest non-empty subset of $\mathbb{N}_0^d$ satisfying the following property: if $x \in \mathcal{S}$ and $\lambda_k(x) > 0$ for some $k = 1, \ldots, K$, then $x + \zeta_k \in \mathcal{S}$. Observe that if $x_0 \in \mathcal{S}$ then we also have that $X_{x_0}(t) \in \mathcal{S}$ for all $t \geq 0$. Hence $\mathcal{S}$ can be taken to be the state space of all the Markov processes described as above with the initial state in $\mathcal{S}$.

Let $\mathcal{P}(\mathcal{S})$ denote the space of probability distributions over $\mathcal{S}$, endowed with the weak topology (see [10]). A set $\mathcal{K} \subset \mathcal{P}(\mathcal{S})$ is compact in this topology if for any $\epsilon > 0$ there is a finite set $A_\epsilon \subset \mathcal{S}$ such that

$$\sup_{\mu\in\mathcal{K}} \mu(A_\epsilon) \geq 1 - \epsilon.$$

For any $x, y \in \mathcal{S}$ let $p_x(t, y)$ denote the following probability

$$p_x(t, y) = \mathbb{P}\left(X_x(t) = y\right). \tag{2}$$

Note that

$$p_x(0, y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Defining $p_x(t, A) = \sum_{y\in A} p_x(t, y)$ for any $A \subset \mathcal{S}$, we can view $p_x(t)$ as an element in $\mathcal{P}(\mathcal{S})$. In fact, $p_x(t)$ is the distribution at time t of the Markov process $(X_x(t))_{t\geq 0}$. The dynamics of $p_x(t)$ is given by Kolmogorov's forward equation which is also called the Chemical Master Equation in the literature. It has the following form:

$$\frac{dp_x(t, y)}{dt} = \sum_{k=1}^{K} p_x(t, y - \zeta_k)\lambda_k(y - \zeta_k) - p_x(t, y)\sum_{k=1}^{K} \lambda_k(y). \tag{4}$$

Theoretically, one can find $p_x(t, y)$ for any $t \geq 0$ and $y \in \mathcal{S}$, by solving this system with initial condition (3). However this system consists of as many ordinary differential equations as the number of elements in $\mathcal{S}$. This system can be explicitly solved if $\mathcal{S}$ is finite, but this only happens in very restrictive cases where all the reactions preserve some conservation relation. Typically, $\mathcal{S}$ is infinite and solving this system analytically or even numerically is quite difficult, if not impossible. Our results are not very useful when $\mathcal{S}$ is finite. For most problems of practical interest $\mathcal{S}$ is infinite and this is what we assume from now on.

The above discussion shows that at the level of distributions, we can view the stochastic dynamics $(X_{x_0}(t))_{t\geq 0}$ as the deterministic dynamics $(p_{x_0}(t))_{t\geq 0}$, which satisfies the system (4). However, the major difficulty in analyzing this deterministic dynamics is that it occurs over an infinite dimensional space $\mathcal{P}(\mathcal{S})$. Nevertheless we can recast the conditions **DC1**, **DC2** and **DC3** in the stochastic setting as below.

SC1 For any x_0, there is a compact set $\mathcal{K}(x_0) \subset \mathcal{P}(\mathcal{S})$ such that $p_{x_0}(t) \in \mathcal{K}(x_0)$ for all $t \geq 0$.

SC2 There exists a compact set $\mathcal{K}_0 \subset \mathcal{P}(\mathcal{S})$ such that for any $x_0 \in \mathcal{S}$ we have $p_{x_0}(t) \in \mathcal{K}_0$ for large values of t.

SC3 There is a $\pi \in \mathcal{P}(\mathcal{S})$ such that for any x_0 we have $p_{x_0}(t) \to \pi$ as $t \to \infty$.

Each of the above conditions give an important insight about the long-term behavior and stability of the stochastic dynamics. The first condition, **SC1**, says that for every $\epsilon > 0$ we can find a finite set $A_\epsilon \subset \mathcal{S}$ such that each $p_{x_0}(t)$ puts at least $(1-\epsilon)$ of its mass in A_ϵ. In other words, for any $t \geq 0$ the probability that $X_{x_0}(t)$ lies outside A_ϵ is less than ϵ. This also means that the Markov process representing the reaction network remains bounded with time, which should be true for most realistic models. If condition **SC2** holds then the evolution of distributions have a compact attractor set in $\mathcal{P}(\mathcal{S})$, where all the trajectories eventually lie irrespective of their starting point. This suggests that in the long run, the family of processes $\{(X_{x_0}(t))_{t\geq 0} : x_0 \in \mathcal{S}\}$, spend most of their time on the same set of states. The last condition **SC3** says that the evolution of distributions have a globally attracting fixed point π. If this holds, then the Markov process representing the reaction dynamics is *ergodic* with π as the unique stationary distribution. When it comes to understanding the long-term behavior of a stochastic process, then ergodicity is a desirable property to have. In the long-run, the proportion of time spent by any trajectory of an ergodic process, in any subset of the state space is equal to the stationary probability of that subset (see (11)). In other words, information about the stationary distribution can be obtained by observing just one trajectory for a sufficiently long time. Such a result is very important in applications. For example, consider a culture with a large number of identical cells with each cell having the same reaction network. If we can show that this intracellular network is ergodic, then by observing the long-term reaction dynamics in a single cell we can obtain statistical information about all the cells at stationarity.

In this paper we develop a general framework for checking conditions **SC1**, **SC2** and **SC3**. However the scope of our paper is broader than that. We obtain easily computable bounds for the statistical moments of the underlying Markov process and investigate when these moments converge with time. We also present conditions for the distribution of the process to be *light-tailed*. The significance of these results is mentioned in the next section.

3 Preliminaries

In this section we discuss the main results of our paper. In particular, we explain how conditions **SC1**, **SC2** and **SC3** can be verified without having to simulate the trajectories of the Markov process representing the reaction dynamics. Intuitively, these conditions can only hold if the Markov process has a low probability of hitting states that have a very large *size*. In our case, the states are vectors in $\mathbb{R}^d$ and so we can measure their *size* by using any norm on $\mathbb{R}^d$. The central theme of this paper is to demonstrate that for many networks, one can analyze the long-term behavior of the reaction dynamics by choosing the *right* norm to measure the state sizes. This *right* norm has the form

$$\|x\|_v = \sum_{i=1}^{d} v_i |x_i|, \tag{5}$$

where v is a positive vector in $\mathbb{R}^d$ satisfying certain conditions based on the reaction network. We specify these conditions in the next section. In the subsequent section we show how the vector v can be determined for a large class of networks by solving a suitably constructed Linear Programming Problem. From now on, we refer to $\|\cdot\|_v$ as the *v-norm* on $\mathbb{R}^d$.

For any positive integer r, let $m^r_{x_0}(t)$ be the r-th moment of $\|X_{x_0}(t)\|_v$ defined by

$$m^r_{x_0}(t) = \mathbb{E}\left(\|X_{x_0}(t)\|^r_v\right) = \sum_{y\in\mathcal{S}} \|y\|^r_v p_{x_0}(t,y). \tag{6}$$

Using Markov's inequality (see [20]) one can show that condition **SC1** holds if for some $r > 0$ and some constant $C_r(x_0)$ we have

$$\sup_{t\geq 0} m^r_{x_0}(t) \leq C_r(x_0). \tag{7}$$

Similarly, condition **SC2** holds if for some $r > 0$ there exists a constant $\widehat{C}_r$ such that

$$\limsup_{t\to\infty} m^r_{x_0}(t) \leq \widehat{C}_r \text{ for all } x_0 \in \mathcal{S}. \tag{8}$$

Relations (7) and (8) give uniform and asymptotic upper-bounds for the r-th moment of the process $(\|X_{x_0}(t)\|_v)_{t\geq 0}$. Using these relations we can also obtain uniform and asymptotic moment bounds for the Markov process $(X_{x_0}(t))_{t\geq 0}$ representing the reaction dynamics (see Corollary 4.3). Such moment bound results have applications in queuing theory and control theory (see [29]). In Theorem 4.2 we show that for certain values of r, (7) and (8) hold and the upper bounds can be easily computed.

Instead of the r-th moment of the process $(\|X_{x_0}(t)\|_v)_{t\geq 0}$, one can ask if the exponential moment of this process is uniformly bounded from above. This will happen if for some $\gamma > 0$ we have

$$\sup_{t\geq 0} \mathbb{E}\left(e^{\gamma\|X_{x_0}(t)\|_v}\right) = \sup_{t\geq 0} \sum_{y\in\mathcal{S}} e^{\gamma\|y\|_v} p_{x_0}(t,y) < \infty. \tag{9}$$

If (9) holds, then for each $t \geq 0$ the distribution $p_{x_0}(t)$ is *light-tailed*, in the sense that its tails are majorized by an exponential decay. It shows that all the cumulants of the distribution $p_{x_0}(t)$ exist, which is an important result for the following reason. There is a considerable body of research dedicated to estimating the moments of the process $(X_{x_0}(t))_{t\geq 0}$ directly without computing the distribution functions $p_{x_0}(t)$. For any integer $r > 0$, one can easily write the differential equations for the dynamics of the first r moments. However when the reaction network has nonlinear interactions, this system of equations is not closed for any r. Various moment closure methods (see [16]) exist that specify ways to close these equations artificially by estimating the moments approximately. One such moment closure method is the cumulant-neglect method which ignores the higher order cumulants of the distribution $p_{x_0}(t)$ for all $t \geq 0$. Of course this method is only valid when

the higher order cumulants exist. This is guaranteed if (9) holds. In Proposition 4.4 we give conditions for verifying (9).

We now come to the question of checking condition **SC3** which says that the process $(X_{x_0}(t))_{t\geq 0}$ is ergodic. This can only happen if the state space $\mathcal{S}$ is *irreducible*, which means that all the states are accessible from each other. Recall the definition of $p_x(t,y)$ from (2). Mathematically, we say that $\mathcal{S}$ is irreducible if for all $x, y \in \mathcal{S}$, we have $p_x(t_1, y) > 0$ and $p_y(t_2, x) > 0$ for some $t_1, t_2 > 0$. In order to check the irreducibility of $\mathcal{S}$, one has to verify that there is no proper subset $\mathcal{S}_1 \subset \mathcal{S}$, such that once the process reaches a state in $\mathcal{S}_1$, it stays in $\mathcal{S}_1$ forever. For reaction networks, this can often be easily checked from the stoichiometry of the reactions. Assuming irreducibility, the ergodicity of the process $(X_{x_0}(t))_{t\geq 0}$ can be established by showing the existence of a *norm-like*[1] function $V : \mathcal{S} \to [0, \infty)$ such that the process $(V(X_{x_0}(t)))_{t\geq 0}$ has the tendency to decrease whenever it reaches a large value (see [26]). In our case, $V(x) = \|x\|_v$ plays the role of this norm-like function and this allows us to verify ergodicity. The precise statement is given in Proposition 4.5.

Suppose that condition **SC3** is satisfied and the process $(X_{x_0}(t))_{t\geq 0}$ is ergodic with stationary distribution π. Using Theorem 4.2 we can identify functions $f : \mathcal{S} \to \mathbb{R}$ for which $\sum_{y\in\mathcal{S}} f(y)\pi(y)$ is finite and

$$\lim_{t\to\infty} \mathbb{E}(f(X_{x_0}(t))) = \sum_{y\in\mathcal{S}} f(y)\pi(y) \tag{10}$$

holds for any $x_0 \in \mathcal{S}$ (see Proposition 4.6). If f is such a function, then the ergodic theorem for Markov processes (see [27]) says that

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t f(X_{x_0}(s))ds = \sum_{y\in\mathcal{S}} f(y)\pi(y) \text{ almost surely,} \tag{11}$$

for any $x_0 \in \mathcal{S}$. From (10), we can deduce which moments of our underlying Markov process converge to their stationary value as time goes to infinity (see Corollary 4.7). Lastly, we also obtain conditions to check if the stationary distribution π is *light-tailed* (see Proposition 4.8).

4 General Results

In this section, we formally present the main results of our paper. Let $\langle\cdot,\cdot\rangle$ be the standard inner product on $\mathbb{R}^d$. From now on we make the following assumption.

Assumption 4.1 *There is a positive vector $v \in \mathbb{R}^d$ along with positive constants c_1, c_2, c_3, c_4, c_5 such that the following conditions hold for all $x \in \mathcal{S}$*

$$\sum_{k=1}^{K} \lambda_k(x)\langle v, \zeta_k\rangle \leq c_1 - c_2\langle v, x\rangle \tag{12a}$$

$$\text{and } \sum_{k=1}^{K} \lambda_k(x)\langle v, \zeta_k\rangle^2 \leq c_3 + c_4\langle v, x\rangle + c_5\langle v, x\rangle^2. \tag{12b}$$

[1] A postive function $V : \mathcal{S} \to \mathbb{R}$ is called *norm-like* if the set $\{x \in \mathcal{S} : V(x) \leq c\}$ is compact for any $c > 0$.

Given a reaction network, the procedure to find such a v is described in the next section. We now explain how this vector v is used. Let $(X_{x_0}(t))_{t\geq 0}$ be the Markov process representing the reaction dynamics. Corresponding to it we can define a *one-dimensional* process $(Y(t))_{t\geq 0}$ as

$$Y(t) = \|X_{x_0}(t)\|_v = \langle v, X_{x_0}(t)\rangle, \tag{13}$$

where the last equality holds because $X_{x_0}(t)$ and v are positive vectors. In this paper we study the process $(X_{x_0}(t))_{t\geq 0}$ by looking at the process $(Y(t))_{t\geq 0}$, which is simpler to study even though it is not a Markov process by itself. The dynamics of $(Y(t))_{t\geq 0}$ has two components *drift* and *diffusion* which have the form $\sum_{k=1}^{K} \lambda_k(x)\langle v, \zeta_k\rangle$ and $\sum_{k=1}^{K} \lambda_k(x)\langle v, \zeta_k\rangle^2$ respectively when $X(t) = x$. Hence (12) gives upper-bounds for the magnitude of these two components. Observe that when the process $(Y(t))_{t\geq 0}$ goes above c_1/c_2 then it experiences a negative drift, suggesting that it will move downwards. This observation is crucial for our analysis.

We are now ready to state the main results of the paper. Their proofs are given in the *Supplementary Material*. In all our results, we assume that we are given a reaction network for which a positive vector v satisfying Assumption 4.1 has been found.

4.1 Moment bounds

Our first result establishes that for certain values of r, relations (7) and (8) hold. Let $r_{\max}$ be the positive number given by

$$r_{\max} = \begin{cases} 1 + \frac{2c_2}{c_5} & \text{if } c_5 > 0 \\ \infty & \text{if } c_5 = 0, \end{cases} \tag{14}$$

where the constants c_2 and c_5 are as in Assumption 4.1.

Theorem 4.2 *For any positive integer r and $x_0 \in \mathcal{S}$ let $m^r_{x_0}(t)$ be defined by (6). If $r < r_{\max}$ then there exist positive constants $C_r(x_0)$ and $\widehat{C}_r$ such that (7) and (8) hold.*

The values of the constants $C_r(x_0)$ and $\widehat{C}_r$ can be explicitly computed using a recursive relationship given in the *Supplementary Material*. For any positive integer r, let $\Psi^r(x_0, t)$ denote the r-th moment of the process $(X_{x_0}(t))_{t\geq 0}$ at time t. Then $\Psi^r(x_0, t)$ is a tensor of rank r whose entry at index $i_1 \dots i_r$ is given by

$$\begin{aligned}\Psi^r_{i_1\dots i_r}(x_0, t) &= \mathbb{E}\left(X_{i_1}(t)X_{i_2}(t)\dots X_{i_r}(t)\right) \\ &= \sum_{y\in\mathcal{S}} y_{i_1}\dots y_{i_r}p_{x_0}(t, y)\end{aligned}$$

for any $i_1, \dots, i_r \in \{1, 2, \dots, d\}$, where $X_{x_0}(t) = (X_1(t), \dots, X_d(t))$, $y = (y_1, \dots, y_d)$ and $p_{x_0}(t)$ is the distribution of $X_{x_0}(t)$. Note that if $v = (v_1, \dots, v_d)$ then $X_i(t) \leq \|X_{x_0}(t)\|_v/v_i$ for any i. This gives us the following corollary of Theorem 4.2.

Corollary 4.3 (Moment Bounds) *Let $\Psi^r(x_0, t)$ be the r-th moment of the process $(X_{x_0}(t))_{t\geq 0}$ at time t. If $r < r_{\max}$ then there exist positive constants $C_r(x_0)$ and*

$\widehat{C}_r$ *such that for any* $i_1, \ldots, i_r \in \{1, 2, \ldots, d\}$ *we have*

$$\sup_{t\geq 0} \Psi^r_{i_1\ldots i_r}(x_0, t) \leq \frac{C_r(x_0)}{\prod_{j=1}^r v_{i_j}}$$

$$\text{and } \limsup_{t\to\infty} \Psi^r_{i_1\ldots i_r}(x_0, t) \leq \frac{\widehat{C}_r}{\prod_{j=1}^r v_{i_j}} \text{ for all } x_0 \in \mathcal{S}.$$

This corollary gives uniform and asymptotic moment bounds for the r-th moment of the Markov process $(X_{x_0}(t))_{t\geq 0}$. Observe that if $c_5 = 0$ then $r_{\max} = \infty$. In this case, Theorem 4.2 says that for each positive integer r and $x_0 \in \mathcal{S}$ there exists a constant $C_r(x_0)$ such that (7) holds. By showing that we have a $C > 0$ such that $C_r(x_0) \leq r!C^r$ for all integers $r \geq 0$, we obtain our next result.

Proposition 4.4 *Suppose that Assumption 4.1 holds with* $c_5 = 0$. *Then there exists a* $\gamma > 0$ *such that*

$$\sup_{t\geq 0} \mathbb{E}\left(e^{\gamma\|X_{x_0}(t)\|_v}\right) = \sup_{t\geq 0} \sum_{y\in\mathcal{S}} e^{\gamma\|y\|_v} p_{x_0}(t, y) < \infty.$$

This proposition gives sufficient conditions for checking that the distribution $p_{x_0}(t)$ is *light-tailed* for each $t \geq 0$.

4.2 Ergodicity and Moment Convergence

The next result verifies the ergodicity of a reaction network satisfying Assumption 4.1. It follows from Theorem 7.1 in Meyn and Tweedie [26].

Proposition 4.5 (Ergodicity) *Let* $(X_{x_0}(t))_{t\geq 0}$ *be the Markov process representing the reaction dynamics and assume that the state space* $\mathcal{S}$ *is irreducible. Then this process is exponentially ergodic in the sense that there exists a unique distribution* $\pi \in \mathcal{P}(\mathcal{S})$ *along with constants* $B, c > 0$ *such that for any* $x_0 \in \mathcal{S}$

$$\sup_{A\subset\mathcal{S}} |p_{x_0}(t, A) - \pi(A)| \leq Be^{-ct} \text{ for all } t \geq 0.$$

This result says that as $t \to \infty$, the distribution $p_{x_0}(t)$ converges to π exponentially fast. Here π is the unique stationary distribution for the dynamics. Henceforth we assume that the process $(X_{x_0}(t))_{t\geq 0}$ is ergodic with stationary distribution π. Using Theorem 4.2 we can prove the following.

Proposition 4.6 *Let* $f : \mathcal{S} \to \mathbb{R}$ *be a function such that for some positive integer* $r < (r_{\max} - 1)$, *there exists a* $C > 0$ *satisfying*

$$|f(x)| \leq C(1 + \|x\|_v^r) \text{ for all } x \in \mathcal{S}.$$

Then $\sum_{y\in\mathcal{S}} f(y)\pi(y)$ *is finite and the relations* (10), (11) *hold.*

For any positive integer r, let Π^r denote the r-th moment of the stationary distribution π. Then Π^r is a tensor of rank r whose entry at index $i_1 \ldots i_r$ is given by

$$\Pi^r_{i_1\ldots i_r} = \sum_{y\in\mathcal{S}} y_{i_1} \ldots y_{i_r} \pi(y)$$

for any $i_1, \ldots, i_r \in \{1, 2, \ldots, d\}$. Proposition 4.6 allows us to prove convergence of moments as time tends to infinity.

Corollary 4.7 *Let* $\Psi^r(x_0, t)$ *be the* r*-th moment of the process* $(X_{x_0}(t))_{t\geq 0}$ *at time* t. *If* $r < (r_{\max} - 1)$ *then* $\Psi^r(x_0, t) \to \Pi^r$ *as* $t \to \infty$.

Another consequence of Proposition 4.6 is that for any $x_0 \in \mathcal{S}$ and positive integer $r < (r_{\max} - 1)$, the quantity $m^r_{x_0}(t)$ (defined by (6)) converges as $t \to \infty$ to $\sum_{y\in\mathcal{S}} \|y\|_v^r \pi(y)$. Then Theorem 4.2 implies that there exists a positive constant $\widehat{C}_r$ such that

$$\sum_{y\in\mathcal{S}} \|y\|_v^r \pi(y) \leq \widehat{C}_r. \tag{15}$$

This relation can be used to obtain moment bounds for the stationary distribution. Note that if $c_5 = 0$ then $r_{\max} = \infty$ and (15) holds for each r. By proving the existence of a constant $C > 0$ such that $\widehat{C}_r \leq r!C^r$ for all positive integers r we get our last result which shows that the stationary distribution is *light-tailed*.

Proposition 4.8 *Suppose that Assumption 4.1 holds with* $c_5 = 0$. *Then there exists a* $\gamma > 0$ *such that*

$$\sum_{y\in\mathcal{S}} e^{\gamma\|y\|_v} \pi(y) < \infty.$$

The framework described above is very general and can be applied to any type of networks that satisfy Assumption 4.1. Below, we specialize the results for two wide classes of networks with mass-action kinetics, namely reaction networks with monomolecular and bimolecular reactions. It will be, however, pointed out in the the examples that the scope is much broader since more general propensities, such as those involving Hill functions, can be considered.

5 Results for affine stochastic reaction networks

Using the analysis tools developed in the previous sections, several very general and broad results can be stated for the class of affine reaction networks, which is the most simple class that can be considered. These networks are, in some sense, analogues of linear dynamical systems. Let us consider an affine reaction network which involves d species that interact through K reaction channels taken among the family

$$\begin{aligned} \emptyset &\xrightarrow{k_0^i} \boldsymbol{S_i} \\ \boldsymbol{S_i} &\xrightarrow{k_i^0} \emptyset \\ \boldsymbol{S_i} &\xrightarrow{k_i^k} \textstyle\sum_{j=1}^d \nu_j^{ik} \boldsymbol{S_j} \end{aligned} \tag{16}$$

where $i = 1, \ldots, d$, $k \in \mathbb{N}$ and $\nu_j^{ik} \in \mathbb{N}_0$. The reaction rates k_0^i, k_i^0 and k_i^k are positive integers. In accordance with (4), the reactions are indexed from $n = 1$ to K, and corresponding propensities and stoichiometries are denoted by $\lambda_n(x)$ and ζ_n, respectively. In the following, it is tacitly assumed that the state-space of the underlying Markov process describing the stochastic chemical reaction network (16) is irreducible.

5.1 Theoretical results

Let us start with several theoretical results that characterize the long-term behavior of affine networks of the form (16).

Theorem 5.1 (Nominal ergodicity) *Let us consider the general affine reaction network* (16) *and assume the state-space of the underlying Markov process is irreducible. Let the matrices $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d_{\geq 0}$ be further defined as*

$$\sum_{n=1}^{K} \lambda_n(x)\langle v, \zeta_n\rangle = x^\intercal A v + b^\intercal v. \tag{17}$$

Then, the following statements are equivalent:

1. *The matrix A is Hurwitz, i.e. all its eigenvalues lie in the open left half-plane.*
2. *There exists a positive vector $v \in \mathbb{R}^d$ such that $Av < 0$.*
3. *The stochastic reaction network has all its moments bounded and converging.*

Moreover, when one of the above statements holds, the Markov process describing the reaction network is exponentially ergodic. ◇

The above result shows that, for affine networks, ergodicity and the existence of moment bounds can be directly inferred from the properties of the matrix A defined in (17). Checking directly whether the matrix A is Hurwitz is indeed an easy problem for which efficient numerical techniques exist. The second statement, on the other hand, characterizes Hurwitzness of A in an implicit way, through the existence of a positive vector $v > 0$. This statement may seem superfluous, at first sight, since simply computing the eigenvalues of A appears to be sufficient for our problem. It turns out, however, that it is the most important one since finding $v > 0$ such that $Av < 0$ is a linear programming problem [4]. A formulation that is much more flexible than the eigenvalue-based one, especially whenever constraints on v needs to be taken into account, as this will be the case when analyzing both affine and quadratic networks in the rest of the paper.

Example 5.2 *Let us consider the following reaction network*

$$\begin{array}{llll} \emptyset & \xrightarrow{k_i} & \boldsymbol{S_i}, & \boldsymbol{S_i} \xrightarrow{\gamma_i} \emptyset \\ \boldsymbol{S_1} & \xrightarrow{k_{12}} & \boldsymbol{S_1}+\boldsymbol{S_2}, & \boldsymbol{S_2} \xrightarrow{k_{21}} \boldsymbol{S_2}+\boldsymbol{S_1} \end{array} \tag{18}$$

where $i = 1, 2$. Then, the moments go unbounded if and only if $\gamma_1\gamma_2 - k_{12}k_{21} \leq 0$ since the matrix A is not Hurwitz in this case.

We extend now Theorem 5.1 to the more pragmatic case of poorly known networks. We therefore assume, in what follows, that the structure of network (the reactions and stoichiometries) is exactly known, but that the reaction rates are subject to uncertainties. This kind of uncertainty reflects naturally into the following parameter-dependent matrices:

$$A(\delta) = A_0 + \sum_{i=1}^{\eta} \delta_i A_i \text{ and } b(\delta) = b_0 + \sum_{i=1}^{\eta} \delta_i b_i \tag{19}$$

where $\delta \in [-1, 1]^\eta$, $\eta \leq K$, are normalized time-invariant, uncertain and independent parameters. The matrices A_0 and b_0 are the nominal matrices which contain the nominal values of the rates. The other matrices represent the uncertainty radius about the nominal values of the rates. We assume further that $A(\delta)$ and $b(\delta)$ do not share common parameters and that the matrices

$$A_+ := \sup_{\delta\in[-1,1]^\eta} \{A(\delta)\} \text{ and } b_+ := \sup_{\delta\in[-1,1]^\eta} \{b(\delta)\} \tag{20}$$

are well-defined in the componentwise sense[2]. This means that $A_+ \geq A(\delta)$ for all $\delta \in [-1, 1]^\eta$ componentwise, and that there exists a $\delta^* \in [-1, 1]^\eta$ for which we have $A_+ = A(\delta^*)$. When such matrices exist, Theorem 5.1 admits the following robustification:

Theorem 5.3 (Robust ergodicity) *Let us consider the general affine reaction network* (16) *described by the uncertain matrices* (19) *that we assume to the admit upper-bounds A_+ and b_+ defined in* (20)*. Assume further that the state-space of the underlying Markov process is irreducible for all uncertain parameter values $\delta \in [-1, 1]^\eta$. Then, the following statements are equivalent:*

1. *The matrix $A(\delta)$ is Hurwitz for all $\delta \in [-1, 1]^\eta$.*
2. *The matrix A_+ is Hurwitz.*
3. *There exists a positive vector $v \in \mathbb{R}^d$ such that $A_+v < 0$.*
4. *The stochastic reaction network* (16)-(19) *has all its moments bounded and converging.*

Moreover, when one of the above statements holds, the Markov process describing the reaction network is is robustly exponentially ergodic. ◇

The important point here is that, whenever the matrices A_+ and b_+ exist, *checking ergodicity of a family of networks is not more complicated than checking ergodicity of a single network*. Note, however, that such a matrix A_+ may not exist due to the presence of coupled entries in the matrix A. This is, for instance, the case when reactions of the type $X_i \longrightarrow X_j$ are involved. In such a case, the above result can be conservatively modified to consider this particular case.

5.2 Numerical results

Several numerical results accompanying the theoretical results of the previous section are described in the following. Many computational results can be extracted from the previous sections, however, only ergodicity verification and first-order moment bounds computation are addressed. The asymptotic first-order moment bound, defined in Theorem 4.2, is given by $\widehat{C}_1 = c_1/c_2$. So the question is, what is the smallest value for such a ratio? Or, in other words, what is the smallest compact set that eventually contains the first-order moments of $\langle v, X(t)\rangle$? Several numerical methods will be discussed to solve, exactly or approximately, this problem.

The first result is a purely algebraic result which gives the optimal (minimal) ratio $\widehat{C}_1^*$:

[2]Given two real matrices A and B with identical dimensions, then $A \geq B$ if and only if $a_{ij} \geq b_{ij}$ for all i, j.

Theorem 5.4 *Let us consider the general affine reaction network* (16) *and assume that the matrix A is irreducible*[3]*. Then, we have that*

$$\widehat{C}_1^* = \frac{b^\intercal v_F(A)}{\lambda_F(A)}$$

where $\lambda_F(A)$ and $v_F(A)$ are the Frobenius eigenvalue and the Frobenius right-eigenvector of A, respectively.

When the matrix A is reducible, the Frobenius eigenvector $v_F(A)$ may contain 0 entries, which would violate the conditions for ergodicity that require a positive v. Additionally, when v contains entries that are equal to 0, the set $\{x \in \mathbb{N}_0^d : \langle v, x\rangle \leq c_1/c_2\}$ is not compact, an irrelevant result when we are interested in finding an attractive compact set for the first-order moments.

The following optimization problem, *fully equivalent to Theorem 5.1*, can be used in order to deal with reducible matrices, by imposing constraints on v:

Optimization problem 5.5 *Let us consider the general affine reaction network* (16) *and assume that the optimization problem*

$$\begin{aligned} &\max_{z,v} z \\ &\quad s.t. \qquad z > 0, v > \varepsilon \\ &\qquad\quad\;\; (zI + A)v \leq 0 \end{aligned} \tag{21}$$

is feasible, with (z^*, v^*) as minimizer, for some $\varepsilon \in \mathbb{R}^d_{>0}$. Then, we have

$$\widehat{C}_1^* \leq b^\intercal v^*/z^* \tag{22}$$

and Theorem 5.1 holds.

Since the feasibility of the above optimization problem is equivalent to Theorem 5.1, this means that when the optimization problem is not feasible (or feasible for some nonpositive z), then all the moments go unbounded.

A striking point concerning the above optimization program is that the numbers of variables and constraints are given by $d+1$ and $2d+1$, respectively. This means that the optimization problem scales *linearly* with respect to the number d of species in the network, and is independent of the number of reactions K. Therefore, from the point of view of this optimization problem, the size of an affine network can be assimilated to the number of species, not the number of reactions. Despite being nonlinear, the above optimization problem can be efficiently solved using a bisection algorithm over z that is globally and geometrically converging to z^*. Each iteration consists of solving a linear program, a class of optimization problems known to be very tractable, and for which numerous advanced solvers exist [4]. These properties, all together, make the overall approach highly scalable, an indispensable property when having the ambition of considering very large real-world networks.

Following (20), the optimization problem (21) admits the following extension to uncertain networks:

[3]A matrix A is irreducible if it is not similar via a permutation to a block upper triangular matrix. Note that this has absolutely no connection with the irreducibility of the state-space of the Markov process describing the network.

Optimization problem 5.6 *Let us consider the general affine reaction network* (16) *with uncertain rate matrices* (19) *which admit the upper-bounds A_+ and b_+ defined in* (20)*. Assume, further, that the optimization problem*

$$\begin{aligned} &\max_{z,v} z \\ &\quad s.t. \qquad z > 0, v > \varepsilon \\ &\qquad\quad\;\; (zI + A_+)v \leq 0 \end{aligned} \tag{23}$$

is feasible, with (z^*, v^*) as minimizer, for some $\varepsilon \in \mathbb{R}^d_{>0}$. Then, we have

$$\widehat{C}_1(\delta) \leq \widehat{C}_1^* \leq b_+^\intercal v^*/z^* \tag{24}$$

for all $\delta \in [-1,1]^\eta$ and Theorem 5.3 holds.

Note that when the matrix A_+ is irreducible, then the Perron-Frobenius Theorem can be applied to get an optimal $\widehat{C}_1^*$.

6 Results for quadratic stochastic reaction networks

Similar results are now presented for quadratic stochastic reaction networks which, in addition to the affine reactions (16), also involve quadratic reactions of the form:

$$\begin{aligned} \boldsymbol{S_i} + \boldsymbol{S_j} &\xrightarrow{k_{ij}^k} \sum_{\ell=1}^d \nu_\ell^{ijk} \boldsymbol{S_\ell} \\ \boldsymbol{S_i} + \boldsymbol{S_j} &\xrightarrow{k_{ij}^0} \emptyset \end{aligned} \tag{25}$$

defined for $i,j = 1,\ldots,d$, $k \in \mathbb{N}$, $\nu_\ell^{ijk} \in \mathbb{N}_0$ and $\ell = 1,\ldots,d$. The reaction rates k_{ij}^k and k_{ij}^0 are positive integers. As for affine networks, it assumed here that the state-space of the underlying Markov process is irreducible.

6.1 Theoretical results for quadratic networks

When quadratic reaction networks of the form (16)-(25) are considered, the left-hand side of condition (12a) can be expressed as

$$\sum_{i=1}^K \lambda_k(x)\langle v, \zeta_k\rangle = x^\intercal M(v)x + x^\intercal A v + b^\intercal v \tag{26}$$

where $M(v) \in \mathbb{R}^{d\times d}$, $A \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d_{\geq 0}$. The matrix $M(v)$ is a symmetric matrix depending affinely on v. Note that, in this case, the matrix A may not be Metzler, although this will be likely the case for most of the realistic reaction networks.

Let $S := \begin{bmatrix}\zeta_1 & \ldots & \zeta_K\end{bmatrix}$ be the stoichiometry matrix of the quadratic reaction network (16)-(25), and let S_q be the restriction of S to quadratic reactions, only.Define further the set

$$\mathcal{N}_q := \left\{v \in \mathbb{R}^d : \; v > 0, \; v^\intercal S_q = 0\right\}$$

which characterizes the set of all positive vectors v such that $v^\intercal$ lies in the left null-space of S_q. When $v \in \mathcal{N}_q$, the

quadratic term $x^\intercal M(v)x$ in (26) vanishes, and equality (26) reduces to

$$\sum_{i=1}^{K} \lambda_k(x)\langle v, \zeta_k\rangle = x^\intercal Av + b^\intercal v$$

which is exactly the same expression as in the case of affine networks. This means that, with the additional constraint that $v \in \mathcal{N}_q$, all the results derived for affine networks directly apply to quadratic networks as well. Along these lines, we obtain the following result:

Theorem 6.1 (Quadratic networks) *Let us consider the quadratic reaction network of the form* (16)-(25) *which we assume to admit a non-empty* $\mathcal{N}_q$.

Then, if there exists a vector $v \in \mathcal{N}_q$ *such that the inequality* $Av < 0$ *holds, the stochastic quadratic reaction network* (16)-(25) *is ergodic and has all its moments bounded and converging.* ◇

Unlike for affine networks where v is unconstrained, Hurwitzness of A is, henceforth, a necessary condition only for Theorem 6.1 to hold. Therefore, the location of the eigenvalues of A does not contain all the necessary information for concluding on ergodicity of the underlying Markov process governing the reaction network. Optimization-based formulations will play a key role in determining a suitable v that satisfies the conditions above.

It should be, moreover, pointed out that the condition of a non-empty set $\mathcal{N}_q$ is central in the statement of Theorem 6.1. When this set is empty, the result cannot be applied in any way. A necessary and sufficient condition for $\mathcal{N}_q$ to be non-empty is that S_q is rank-deficient, which corresponds to a certain form of rarity for the quadratic reactions. This condition may seem restrictive at first sight, but it will be shown that several important reaction networks from the literature actually fall into this category.

Whenever $\mathcal{N}_q$ is empty, linear programming can still be applied provided that the matrix $M(v)$ exhibits a particular sparsity pattern defined below:

Definition 6.2 *We say that a symmetric real matrix* $Z = [z_{ij}]$ *is of class* $\mathcal{S}_I$ *if the elements of* Z *satisfy the conditions*

$$z_{ii}z_{ij}z_{jj} = 0 \tag{27}$$

for all $i \neq j$. *A network is said to be* $\mathcal{S}_I$*-quadratic if its matrix* $M(v)$, *defined in* (26), *is of class* $\mathcal{S}_I$ *for all* $v \in \mathbb{R}^d_{>0}$.

The underlying reason for considering this class of matrices lies in the fact that, whenever a symmetric real matrix $Z \in \mathbb{R}^{d\times d}$ is of class $\mathcal{S}_I$, checking whether $x^\intercal Zx \leq 0$ for all $x \in \mathbb{R}^d_{\geq 0}$ is equivalent to the condition that $Z \leq 0$ componentwise. Based on this fact, the following result is obtained:

Theorem 6.3 ($\mathcal{S}_I$-Quadratic networks) *Let us consider the quadratic reaction network of the form* (16)-(25) *which we assume to be* $\mathcal{S}_I$*-quadratic. Assume further that there exists a vector* $v > 0$ *such that* $Av < 0$ *and* $M(v) \leq 0$.

Then, the quadratic stochastic reaction network is ergodic and has its moments up to order $\lfloor 1 + 2c_2/c_5 \rfloor - 1$ *bounded and converging.* ◇

6.2 Numerical results for quadratic networks

It is shown here that, once again, the theoretical results can be easily turned into linear programs that can be checked in a very efficient way. The following result is the numerical translation of Theorem 6.1:

Optimization problem 6.4 *Let us consider a quadratic reaction network* (16)-(25) *admitting a nonempty set* $\mathcal{N}_q$. *Assume that the optimization problem*

$$\begin{array}{ll} \max_{z,v} z & \\ s.t. & z > 0, v > \varepsilon \\ & (zI + A)v \leq 0 \\ & v^\intercal S_q = 0. \end{array} \tag{28}$$

is feasible with (z^*, v^*) *as minimizer. Then, we have*

$$\widehat{C}_1^* \leq b^\intercal v^*/z^*,\ v^* \in \mathcal{N}_q \tag{29}$$

and Theorem 6.1 holds.

Note the presence of the additional constraint $v^\intercal S_q = 0$ which imposes to $v^\intercal$ to lie in the left null-space of S_q. We can see, again, that the complexity scales linearly with the number of species since the number of variables and constraints is $d + 1$ and $3d + 1$, respectively. Therefore, quadratic networks admitting a nonempty set $\mathcal{N}_q$ may be, in fact, not more complicated than affine networks.

The following optimization problem is the computational counterpart of Theorem 6.3:

Optimization problem 6.5 *Let us consider a* $\mathcal{S}_I$*-quadratic reaction network of the form* (16)-(25). *Assume further that the nonlinear optimization problem*

$$\begin{array}{ll} \max_{z,v} z & \\ s.t. & z > 0, v > \varepsilon \\ & (zI + A)v \leq 0 \\ & M(v) \leq 0. \end{array} \tag{30}$$

is feasible with (z^*, v^*) *as minimizer. Then, we have*

$$\widehat{C}_1^* \leq b^\intercal v^*/z^* \tag{31}$$

and Theorem 6.3 holds.

The above optimization problem does not scale as nicely as the optimization problem (30) since, in the worst case, the number of constraints related to $M(v)$ grows as $d(d+1)/2$, hence quadratically in terms of the number of species. The number of variables is, however, still equal to $d + 1$. Despite the complexity increase, the problem remains tractable due to the linear program structure.

7 Examples

7.1 Finding an attractive compact set for the first-order moments trajectories

The goal of this section is to describe the computation of compact sets that eventually include the first order moment of $\langle v, X(t)\rangle$. Since we are interested in finding the smallest one (or an approximation of it), the goal is therefore to minimize the ratio c_1/c_2 in (12a), a task which

can be performed using the optimization problems (21) or (28), according to the type of networks that is considered. Due to the moment closure problem [16], analytical expressions for the steady-state values of the moments of quadratic networks are not available. This class of networks is, therefore, the most interesting class to consider. Let us indeed consider the following quadratic network

$$\begin{array}{rcl} \emptyset & \xrightarrow{k} & \boldsymbol{S_1}, \\ \boldsymbol{S_1}+\boldsymbol{S_1} & \xrightarrow{k_{12}} & \boldsymbol{S_2}, \\ \boldsymbol{S_2} & \xrightarrow{\gamma_2} & \emptyset. \end{array} \qquad \begin{array}{rcl} \boldsymbol{S_1} & \xrightarrow{\gamma_1} & \emptyset \\ \boldsymbol{S_2} & \xrightarrow{k_{21}} & \boldsymbol{S_1}+\boldsymbol{S_1} \end{array} \tag{32}$$

representing a dimerization process, i.e. $\boldsymbol{S_1}$ dimerizes in $\boldsymbol{S_2}$. It is easily seen that this network is irreducible since any point in the state-space can be reached from any other point in a finite number of reactions having nonzero propensities. Choosing v in $\mathcal{N}_q$ where $S_q = \begin{bmatrix} -2 & 1 \end{bmatrix}^T$, e.g. $v^\intercal = \begin{bmatrix} 1 & 2 \end{bmatrix}$, theoretical calculations show that condition (12a) holds with $c_1^* = k$ and $c_2^* = \min\{\gamma_1, \gamma_2\}$. This proves that the system is exponentially ergodic and that all the moments are bounded and converge to a unique steady-state value. Solving now the optimization problem (28) with numerical values $k = 1$, $\gamma_1 = \gamma_2 = 0.2$ (the values of k_{12} and k_{21} are unimportant here), we get that $c^* = 5$, which coincides with the theoretical value. To validate the compact set calculation, Monte-Carlo simulations are performed and yield the values $E_1 = 1.17 \pm 0.01$ and $E_2 = 1.927 \pm 0.02$ for the equilibrium values of the average number of species $\boldsymbol{S_1}$ and $\boldsymbol{S_2}$. Evaluating then $\langle v, E \rangle$ with these values yields

$$\langle v, E \rangle = E_1 + 2E_2 = 5.024 \pm 0.05, \tag{33}$$

showing that the compact set agrees very well with the equilibrium values. Note also that, by virtue of Theorem 4.2, we also have $\lim_{t\to\infty} \mathbb{E}[\langle v, X(t) \rangle] \leq c_1^*/c_2^*$.

To illustrate these results, several trajectories for $E[X_1(t)]$ and $\mathbb{E}[X_2(t)]$ for different initial conditions are plotted in Fig. 1 where we can see that all the trajectories converge to a point, inside the compact set (the surface below the dashed line), very close to the boundary. Note, moreover, that the trajectories starting from inside the compact set never cross the boundary of the set, as predicted by the theory.

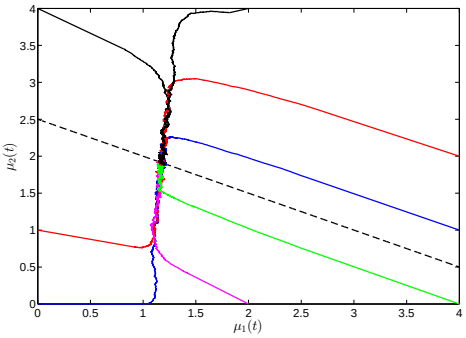

Figure 1: Trajectories of the first order moments $\mu_1(t) = \mathbb{E}[X_1(t)]$ and $\mu_2(t) = \mathbb{E}[X_2(t)]$ of network (32) for different initial conditions (averaging is performed over 5000 cells). The trajectories converge to the unique steady-state value located inside the compact set (the surface below the dashed line), very close to the boundary.

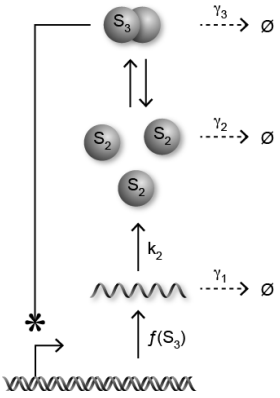

Figure 2: Feedback loop with arbitrary feedback rule.

7.2 Feedback loop

Let us consider the feedback loop network of Fig. 2 which can be represented by the set of reactions

$$\begin{array}{rcl} \boldsymbol{S_1} & \xrightarrow{k_2} & \boldsymbol{S_1}+\boldsymbol{S_2}, \\ \boldsymbol{S_3} & \xrightarrow{k_{32}} & \boldsymbol{S_2}+\boldsymbol{S_2}, \\ \boldsymbol{S_i} & \xrightarrow{\gamma_i} & \emptyset. \end{array} \qquad \begin{array}{rcl} \emptyset & \xrightarrow{f(\boldsymbol{S_3})} & \boldsymbol{S_1} \\ \boldsymbol{S_2}+\boldsymbol{S_2} & \xrightarrow{k_{23}} & \boldsymbol{S_3} \end{array} \tag{34}$$

which represents a gene expression network, with $\boldsymbol{S_1}$ as mRNA and $\boldsymbol{S_2}$ as protein, implementing a feedback loop for controlling its own expression through the protein dimer $\boldsymbol{S_3}$. The function $f(\cdot)$ can either be an activator function (positive feedback) or an inhibitor function (negative feedback), but the results are valid for any bounded function $f(\cdot)$. Choosing $v \in \mathcal{N}_q$ and solving for the conditions of Theorem 6.1 yield that $c_1 = v_1 \sup_{x\geq 0}\{f(x)\}$ and that c_2 can be chosen either as $c_2 = \min\{\gamma_2, \gamma_3\}$ if $\gamma_1 > \min\{\gamma_2, \gamma_3\}$ or can be made arbitrarily close to γ_1, otherwise, through a suitable choice for $v \in \mathcal{N}_q$. We therefore have the following concluding statement:

Result 7.1 *For any values of the rate parameters and any bounded function $f(\cdot)$, the feedback loop with dimerization* (34) *is ergodic and has all its moments bounded and converging.*

The following stronger result generalizes the above result to a feedback loop with multimerization of arbitrary order:

Result 7.2 *For any values of the rate parameters and any bounded function $f(\cdot)$, the feedback loop with arbitrary multimerization and full or partial degradation is ergodic and has all its moments bounded and converging.*

7.3 Stochastic switch

Let us consider the stochastic switch of [33] described by the affine stochastic reaction network

$$\begin{array}{rcl} \emptyset & \xrightarrow{f_1(\boldsymbol{S_2^1})} & \boldsymbol{S_1^0}, \\ \emptyset & \xrightarrow{f_2(\boldsymbol{S_1^1})} & \boldsymbol{S_2^0}, \\ \boldsymbol{S_i^j} & \xrightarrow{\gamma_{i,j}} & \emptyset. \end{array} \qquad \begin{array}{rcl} \boldsymbol{S_1^0} & \xrightarrow{k_1} & \boldsymbol{S_1^0}+\boldsymbol{S_1^1} \\ \boldsymbol{S_2^0} & \xrightarrow{k_2} & \boldsymbol{S_2^0}+\boldsymbol{S_2^1} \end{array} \tag{35}$$

Above $\boldsymbol{S}_i^0$ and $\boldsymbol{S}_i^1$ represent mRNAs and proteins of gene i, respectively. The functions $f_1(\cdot)$ and $f_2(\cdot)$ are defined as inhibiting Hill functions but, as in the example above, any bounded functions are allowed. We can state the following result:

Result 7.3 *For any values of the rate parameters and any bounded functions $f_1(\cdot)$ and $f_2(\cdot)$, the stochastic switch (35) is ergodic and has all its moments bounded and converging.*

As for the feedback loop, the above result generalizes to

Result 7.4 *For any values of the rate parameters and any functions bounded $f_1(\cdot)$ and $f_2(\cdot)$, the stochastic switch with arbitrary multimerization order and full or partial degradation is ergodic and has all its moments bounded and converging.*

7.4 Repressilator

We consider here a stochastic repressilator [7] model involving N genes and negative feedback implemented in terms of Hill functions denoted by $f_i(\cdot)$, where i corresponds to the gene number; see Fig. 3. We have the following result:

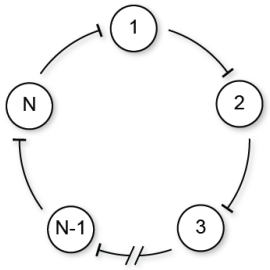

Figure 3: N-gene repressilator.

Result 7.5 *For any values of the rate parameters and any functions $f_i(\cdot)$'s, the stochastic N-gene repressilator is ergodic and has all its moments bounded and converging.*

The result can be generalized to account for arbitrary multimerization of proteins to:

Result 7.6 *For any values of the rate parameters and any functions $f_i(\cdot)$'s, the stochastic N-gene repressilator with arbitrary multimerization is ergodic and has all its moments bounded and converging.*

7.5 Stochastic SIR model

We consider here the following SIR-model, similar to the one in [6], defined as

$$\begin{array}{llllll} \emptyset \xrightarrow{k_s} \boldsymbol{S}, & \emptyset \xrightarrow{k_i} \boldsymbol{I}, & \boldsymbol{S} \xrightarrow{\gamma_s} \emptyset \\ \boldsymbol{I} \xrightarrow{\gamma_i} \emptyset, & \boldsymbol{R} \xrightarrow{\gamma_r} \emptyset, & \boldsymbol{S}+\boldsymbol{I} \xrightarrow{k_{si}} 2\boldsymbol{I} \\ \boldsymbol{I} \xrightarrow{k_{ir}} \boldsymbol{R}, & \boldsymbol{R} \xrightarrow{k_{rs}} \boldsymbol{S}. \end{array} \tag{36}$$

where birth and death reactions represent people entering and leaving the process, respectively. The only quadratic reaction is the contamination reaction which turns one susceptible person into an infectious one. The two last reactions represent how infectious people are recovering and how recovered people become susceptible again.

We then have the following result:

Result 7.7 *For any values of the rate parameters, the SIR-model (36) is ergodic and has all its moments bounded and converging.*

7.6 Circadian clock

Let us consider the circadian oscillator of [34], see Fig. 4, which is a network involving 9 species and 18 reactions. Using the rate parameters values of [34], we obtain the typical oscillatory trajectory depicted in Fig. 5. Applying the developed theory on this model, we obtain the following result:

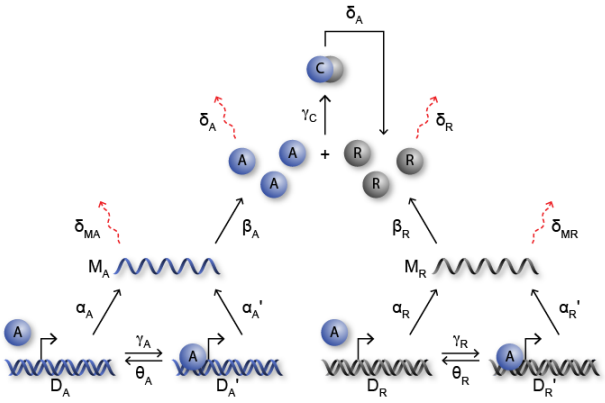

Figure 4: Circadian clock model of [34].

Result 7.8 *For any values of the rate parameters, the circadian clock model of [34] is ergodic and has all its moments bounded and converging.*

Using, for instance, the values of [34] and solving for the optimization problem (28), we find that $c_1 = 402.5768$ and $c_2 = 0.1992$, and we obtain

$$\lim_{t\to\infty} \mathbb{E}[\langle v, X(t)\rangle] \leq 2020.778 \tag{37}$$

for the specific v that has been obtained. When averaging over a population of 2000 cells, we obtain the sample-average trajectories depicted in Fig. 6, where we observe convergence to stationary values, even though the sample-paths are oscillatory. From ergodicity, we can even state that these fixed points for the sample-averages are globally attracting.

7.7 p53 model

Let us us consider one of the oscillatory p53 models of [13], which is described by the reactions

$$\begin{array}{llllllll} \emptyset \xrightarrow{k_1} \boldsymbol{S_1}, & \boldsymbol{S_1} \xrightarrow{k_2} \emptyset, & \boldsymbol{S_1} \xrightarrow{f(\boldsymbol{S_1},\boldsymbol{S_3})} \emptyset \\ \boldsymbol{S_3} \xrightarrow{k_6} \emptyset, & \boldsymbol{S_2} \xrightarrow{k_5} \boldsymbol{S_3}, & \boldsymbol{S_1} \xrightarrow{k_4} \boldsymbol{S_1}+\boldsymbol{S_2}. \end{array} \tag{38}$$

where $\boldsymbol{S_1}$ is the number of p53 molecules, $\boldsymbol{S_2}$ the number of precursor of Mdm2 molecules and $\boldsymbol{S_3}$ the number of molecules of Mdm2. The function $f(x,y) = \dfrac{k_3 y}{x+k_7}$ implements a nonlinear feedback on the degradation rate of p53. We have the following result:

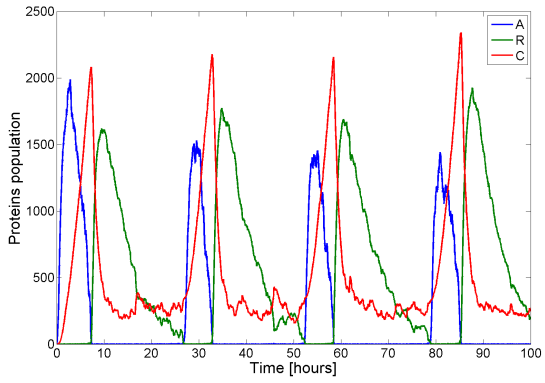

Figure 5: Sample-path of the species of the circadian clock model.

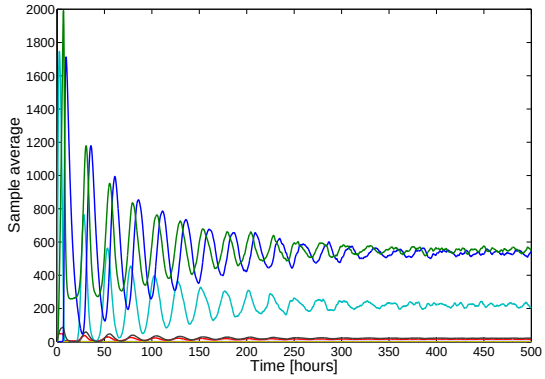

Figure 6: Time evolution of the sample averages of the species of the circadian clock model (2000 cells averaging).

Result 7.9 *For any values of the rate parameters, the oscillatory p53 model* (38) *is ergodic and has all its moments bounded and converging.*

7.8 Lotka-Volterra model

We consider here the stochastic reaction network

$$\begin{array}{rcl} \emptyset & \xrightarrow{\alpha_i} & \boldsymbol{S_i}, \qquad \boldsymbol{S_i} \xrightarrow{\beta_i} \boldsymbol{S_i}+\boldsymbol{S_i} \\ \boldsymbol{S_i}+\boldsymbol{S_j} & \xrightarrow{\gamma_{ij}} & \boldsymbol{S_j}, \qquad \boldsymbol{S_i} \xrightarrow{\delta_i} \emptyset \end{array} \tag{39}$$

which is an open analogue of the deterministic Lotka-Volterra system of [14]. The first set of reactions represent immigration, the second one reproduction, the third one competition due to overpopulation and the last one deaths/migrations. We obtain then the following result, which is a stochastic analogue of the results in [5] obtained in the deterministic setting:

Theorem 7.10 *Let us define* $\Gamma(v) = [v_i\gamma_{ij}]$ *and assume that one of the following assumptions hold:*

1. *there exists* $v > 0$ *such that the matrix* $\Gamma(v) + \Gamma(v)^\intercal$ *is positive definite;*

2. *there exists* $v > 0$ *such that the* $\Gamma(v)+\Gamma(v)^\intercal$ *is copositive*[4] *and* $\beta_i - \delta_i < 0$ *for all* $i = 1, \ldots, n$.

[4] A matrix Z is said to be copositive if $x^T Z x \geq 0$ for all $x \geq 0$.

Then, the stochastic reaction network (39) *is ergodic and all the moments up to order* $\left\lfloor 1 + \frac{2c_2}{c_5} \right\rfloor - 1$ *are bounded and converging.*

Acknowledgments

The authors gratefully thank Stephanie Aoki for her help in drawing pictures.

References

[1] Adam P. Arkin, Christopher V. Rao, and Denise M. Wolf. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231–237, 2002.

[2] Jordi Bascompte. Structure and dynamics of ecological networks. *Science*, 329(5993):765–766, 2010.

[3] Seth I. Berger and Ravi Iyengar. Network analyses in systems pharmacology. *Bioinformatics*, 25(19):2466–2472, 2009.

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, MA, USA, 2004.

[5] N. Champagnat, P.-E. Jabin, and G. Raoul. Convergence to equilibrium in competitive lotka-volterra and chemostat systems. *Compte Rendus de l'Académie des Sciences - Mathématique*, 348(23-24):1267–1272, 2010.

[6] W.-Y. Chen and S. Bokka. Stochastic modeling of nonlinear epidemiology. *J. theor. Biol.*, 234:455–470, 2005.

[7] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[8] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[9] P. Érdi and J. Tóth. *Mathematical models of chemical reactions*. Nonlinear Science: Theory and Applications. Princeton University Press, Princeton, NJ, 1989. Theory and applications of deterministic and stochastic models.

[10] S. N. Ethier and T. G. Kurtz. *Markov processes : Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[11] Martin Feinberg. Complex balancing in general kinetic systems. *Arch. Rational Mech. Anal.*, 49:187–194, 1972/73.

[12] Martin Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Arch. Rational Mech. Anal.*, 132(4):311–370, 1995.

[13] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, P. Polak, G. Lahav, and U. Alon. Oscillations and variability in the p53 system. *Molecular Systems Biology*, 2:2006.0033, 2006.

[14] K. Gopalsamy. Global asymptotic stability in volterra's population systems. *Journal of Mathematical Biology*, 19:157–168, 1984.

[15] John Goutsias. Classical versus stochastic kinetics modeling of biochemical reaction systems. *Biophysical Journal*, 92(7):2350 – 2365, 2007.

[16] J. P. Hespanha. Moment closure for biochemical networks. In *3rd International Symposium on Communications, Control and Signal Processing*, pages 142–147, St. Julian's, Malta, 2008.

[17] H. Hethcote. The mathematics of infectious diseases. *SIAM Review*, 42(4):599–653, 2000.

[18] F. Horn. Necessary and sufficient conditions for complex balancing in chemical kinetics. *Arch. Rational Mech. Anal.*, 49:172–186, 1972/73.

[19] F. Horn and R. Jackson. General mass action kinetics. *Arch. Rational Mech. Anal.*, 47:81–116, 1972.

[20] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[21] Hassan K. Khalil. *Nonlinear systems*. Macmillan Publishing Company, New York, 1992.

[22] Andrzej M. Kierzek, Jolanta Zaim, and Piotr Zielenkiewicz. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *Journal of Biological Chemistry*, 276(11):8165–8172, 2001.

[23] Matthew D. Levin, Carl J. Morton-Firth, Walid N. Abouhamad, Robert B. Bourret, and Dennis Bray. Origins of individual swimming behavior in bacteria. *Biophysical Journal*, 74(1):175 – 181, 1998.

[24] H. H. McAdams and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in genetics : TIG*, 15(2):65–69, February 1999.

[25] Harley H. McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci., Biochemistry*, 94:814–819, 1997.

[26] Sean P. Meyn and R. L. Tweedie. Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.*, 25(3):518–548, 1993.

[27] J. R. Norris. *Markov chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.

[28] Jason A. Papin, Tony Hunter, Bernhard O. Palsson, and Shankar Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol*, 6(2):99–111, 02 2005.

[29] Robin Pemantle and Jeffrey S. Rosenthal. Moment conditions for a sequence with negative drift to be uniformly bounded in lr. *Stochastic Processes and their Applications*, 82(1):143–155, 1999.

[30] Robert Schuetz, Nicola Zamboni, Mattia Zampieri, Matthias Heinemann, and Uwe Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.

[31] S. Smale. *Differential Equations, Dynamical Systems & an introduction to Chaos*. Academic Press, 2004.

[32] Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

[33] T. Tian and K. Burrage. Stochastic models for regulatory networks of the genetic toggle switch. *Proc. Natl. Acad. Sci.*, 103(22):8372–8377, 2006.

[34] J. M. G. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillator. *Proc. Natl. Acad. Sci.*, 99(9):5988–5992, 2002.